PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-041078
(43)Date of publication of application : 08.02.2002

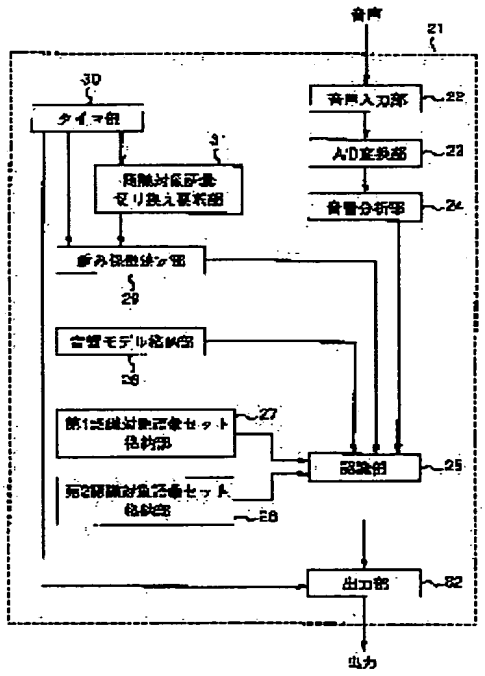
(51)Int.CI. G10L 15/06
G10L 15/10
G10L 15/00
G10L 15/28

(21)Application number : 2000-220576
(22)Date of filing : 21.07.2000

(71)Applicant : SHARP CORP
(72)Inventor : HONDA KAZUMASA
TSURUTA AKIRA
KANZA HIROYUKI

(54) VOICE RECOGNITION EQUIPMENT, VOICE RECOGNITION METHOD AND PROGRAM RECORDING MEDIUM

(57)Abstract:
PROBLEM TO BE SOLVED: To obtain high recognition accuracy even when the recognition object vocabulary is automatically changed.
SOLUTION: A recognition means 25 calculates the likelihood P of the vocabulary which constitutes the recognition object vocabulary sets A, B stored in the first and second recognition object vocabulary storing parts 27, 28 using the acoustics model in an acoustics model storing part 26. The change of the recognition object vocabulary sets accompanied with the display contents in an output part 32 is performed at the time by changing the values of weights w1, w2 which multiply to P between '1' and the appointed value 'a' near 0 (zero) proportioned the passing time from the requested change time t0. As the result, even when the speaker misses the utterance chance and the recognition object vocabulary is changed automatically, high recognition result can be obtained if the speaker utters the recognition object vocabulary before changing because the calculations of the likelihood w × P are also performed.



LEGAL STATUS

[Date of request for examination] 26.07.2002
[Date of sending the examiner's decision of rejection]
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
[Date of final disposal for application]
[Patent number] 3563018

[Date of registration] 11.06.2004

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

先願2

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2002－41078
（P2002－41078A）
(43)公開日　平成14年２月８日(2002.2.8)

(51)Int.Cl.'	識別記号	ＦＩ		テーマコード'(参考)
Ｇ１０Ｌ　15/06		Ｇ１０Ｌ　3/00	５２１Ｖ	５Ｄ０１５
15/10			５３１Ｇ	
15/00			５５１Ｐ	
15/28				

審査請求　未請求　請求項の数６　ＯＬ　（全 11 頁）

(21)出願番号　特願2000－220576(P2000－220576)

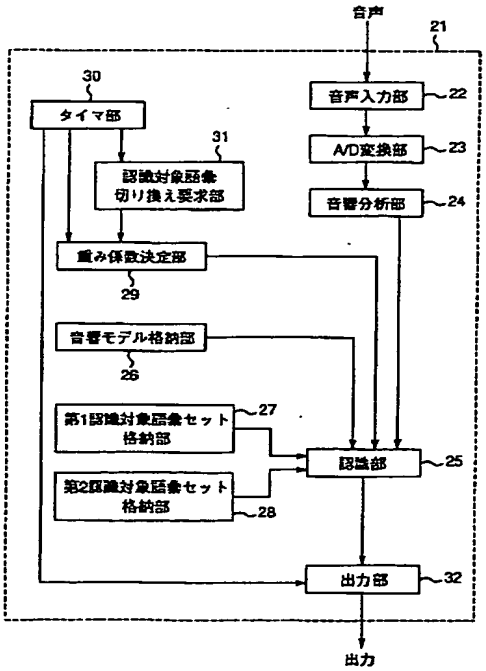
(22)出願日　平成12年７月21日(2000.7.21)

(71)出願人　000005049
シャープ株式会社
大阪府大阪市阿倍野区長池町22番22号
(72)発明者　本田　和正
大阪府大阪市阿倍野区長池町22番22号　シャープ株式会社内
(72)発明者　鶴田　彰
大阪府大阪市阿倍野区長池町22番22号　シャープ株式会社内
(74)代理人　100062144
弁理士　青山　葆　（外１名）

最終頁に続く

(54)【発明の名称】　音声認識装置および音声認識方法、並びに、プログラム記録媒体

(57)【要約】
【課題】　自動的に認識対象語彙を切り換える場合でも高認識精度を得る。
【解決手段】　認識部２５は、音響モデル格納部２６の音響モデルを用いて第１,第２認識対象語彙セット格納部２７に格納された認識対象語彙セットＡ,Ｂを構成する単語の尤度Ｐを算出する。その際に、出力部３２の表示内容の切り換えに伴う認識対象語彙セットの切り換えは、切り換え前後の認識対象語彙セットを構成する単語の尤度Ｐに掛ける重みw_1,w_2の値を切り換え要求時刻ｔ。からの経過時間に比例して「１」と０近傍の所定値「ａ」との間で切り換えることで行う。その結果、話者が認識対象語彙の発声の機会を逸し、且つ、自動的に認識対象語彙が切り換えられても、切り換え前の認識対象語彙セットの単語に関する尤度ｗ・Ｐの計算も行われるために、話者が切り換え前の認識対象語彙で発声しても高い認識結果が得られる。



1

【特許請求の範囲】

【請求項１】　入力された音声を認識する認識部と、この認識部の認識結果を含む情報を出力する出力部と、上記認識時に用いられる認識対象語彙が格納された認識語彙格納部と、タイマ部と、このタイマ部からの時刻信号に基づいて上記認識対象語彙の切り換えを要求する認識対象語彙切り換え要求部を有する音声認識装置において、

上記出力部は、複数の出力内容を切り換え出力するようになっており、

上記認識対象語彙は、上記出力部の出力内容に対応した認識対象語の集合でなる複数の認識対象語彙セットに分類され、上記認識対象語彙の切り換えは上記認識対象語彙セットの単位で行われるようになっており、

上記タイマ部からの時刻信号に基づいて、上記各認識対象語彙セット用の重みを決定する重み決定部を備えて、

上記認識部は、上記全認識対象語彙セットおよび上記決定された各重みを用いて、入力音声を認識するようになっていることを特徴とする音声認識装置。

【請求項２】　請求項１に記載の音声認識装置において、

上記重み決定部は、上記認識対象語彙切り換え要求部によって認識対象語彙の切り換えが要求されてから重み決定までの経過時間に応じて、切り換え前の認識対象語彙セット用の重みを低下させる一方、切り換え後の認識対象語彙セット用の重みを上昇させるようになっていることを特徴とする音声認識装置。

【請求項３】　請求項１あるいは請求項２に記載の音声認識装置において、

上記認識部は、上記全認識対象語彙セットを構成する各語の尤度を算出し、各語の尤度の値に各語が属する認識対象語彙セット用の重みを掛け、その値が最も高い語を認識結果とするようになっていることを特徴とする音声認識装置。

【請求項４】　請求項２に記載の音声認識装置において、

上記出力部は、上記認識対象語彙切り換え要求部からの認識対象語彙切り換え要求がなされた時点に出力している出力内容に対応する認識対象語彙セット用の重みの値と、次に出力すべき出力内容に対応する認識対象語彙セット用の重みの値との差が所定値未満になると、上記出力内容を切り換えるようになっていることを特徴とする音声認識装置。

【請求項５】　入力された音声を認識対象語彙を用いて認識して認識結果を出力するに際して、タイマ部からの時刻信号に基づいて上記認識対象語彙の切り換えを自動的に行う音声認識方法において、

複数の出力内容を出力部に切り換え出力し、

上記各出力内容に対応した認識対象語の集合でなる複数の認識対象語彙セットの単位で、上記認識対象語彙の切り換えを行い、

上記タイマ部からの時刻信号に基づいて、上記各認識対象語彙セット用の重みを決定し、

上記全認識対象語彙セットおよび上記決定された各重みを用いて、上記入力音声の認識を行うことを特徴とする音声認識方法。

【請求項６】　コンピュータを、

請求項１における認識部,出力部,タイマ部,認識対象語彙切り換え要求部および重み決定部として機能させる音声認識処理プログラムが記録されたことを特徴とするコンピュータ読出し可能なプログラム記録媒体。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】この発明は、コンピュータや携帯情報端末に搭載されて人間の発声による音声を認識する音声認識装置および音声認識方法、並びに、音声認識処理プログラムを記録したプログラム記録媒体に関する。

【０００２】

【従来の技術】音声認識装置において、認識精度を高めるために、必要に応じて認識対象語彙を切り換えるという認識方法がある。このような認識方法を用いた音声認識装置の応用例として、パーソナルコンピュータや日本語ワードプロセッサ等の表示装置を有する機器において、表示装置を用いたメニュー表示による機器の操作ガイドを、音声認識を用いて行うことが考えられる。

【０００３】上述のような操作ガイドによれば、操作方法や操作による効果の表示を画面で確認しながら操作を学ぶことができる。そして、上記表示装置の画面が狭い等、上記表示装置からの情報量が少ない場合には、複数の機器操作に関する操作ガイドの表示を時間の経過と共に自動的に切り換える場合がある。このような操作ガイドに音声を用いれば、利用者にとって分り易く、且つ、操作ボタンの数を減らして操作を簡単にすることができる。その場合、複数の機器操作に関する操作ガイドの表示の切り換えと共に認識対象語彙を切り換えれば、認識対象語彙を少なくすることができるので高い認識精度を得ることができる。

【０００４】このような認識対象語彙を切り換える認識方法の応用においては、切り換え表示する各メニューに関連のある認識対象語彙のセットを、メニュー数分だけ複数記憶しておく。そして、利用者の操作や時間の経過等によるメニュー表示の切り換えに同期して認識対象語彙を切り換えることによって、夫々のメニューにおいては必要最小限の語彙を対象に認識処理を行うことができ、認識精度を向上させることができるのである。その場合、時間の経過と共にメニュー表示を自動的に切り換える際には、機器が自動的に認識対象語彙をも切り換えることになる。

【０００５】以下、上記認識対象語彙の切り換えが可能

な音声認識装置について説明する。図４は、上記認識対象語彙の切り換えが可能な音声認識装置の一例を示すブロック図である。ここで、本音声認識装置１は、認識対象語彙の切り換えおよび出力部１３による表示内容の切り換えは、所定時間毎に音声認識装置１自身が自動的に行うものとする。音声認識装置１は、Ａ/Ｄ(アナログ/ディジタル)変換部２、音響分析部３、認識部４、音響モデル格納部５、認識対象語彙格納･判定部６、現認識対象語彙識別子記憶部７、タイマ部８、認識対象語彙切り換え要求部９、認識対象語彙切り換え要求時刻記憶部１０、音声検出部１１、音声時刻記憶部１２、出力部１３で構成される。

【０００６】話者によって上記音声認識装置１に入力された音声は、Ａ/Ｄ変換部２に送出されてディジタル化される。そして、このディジタル化された音声波形は、音響分析部３で、２０msec～４０msecの区間毎に比較的短時間の時間窓を掛けると共に、８msec～１６msec毎に上記時間窓をシフトしていく短時間スペクトル分析の手法によって分析される。上記時間窓によって切り出された音声波形は、切り出し時の時間長を有するフレームと呼ばれる単位の特徴ベクトルの時系列に変換される。ここで、上記特徴ベクトルは、その時刻における音声スペクトルの特徴量を抽出したもので、通常は１０次元～１００次元であり、ＬＰＣ(線形予測分析)メルケプストラム係数等が広く用いられている。こうして変換された特徴ベクトルは認識部４に送出されると共に、音声入力の開始を検出する音声検出部１１にも出力される。そうすると、音声時刻記憶部１２は、音声検出部１１からの音声入力開始信号とタイマ部８からの時刻信号とに基づいて、音声入力の開始時刻を検出して記憶する。

【０００７】上記音響モデル格納部５には、認識単位毎に用意されたＨＭＭ(隠れマルコフモデル)が用意されている。ここで、上記認識単位としては、音素や単語が広く用いられている。また、ＨＭＭとは、複数個の状態を有する非決定性確率有限オートマトンであり、非定常信号源を定常信号源の連結で表す統計的信号源モデルである。尚、出力確率や遷移確率等のパラメータは、対応する学習音声を与えてバウム・ウェルチアルゴリズムと呼ばれるアルゴリズム等によって予め学習されている。以下、音響モデル格納部５には、認識単位が音素であるＨＭＭが記憶されているものとする。

【０００８】上記認識対象語彙の切り換えの動作は、特開平６－３３７６９５号公報に開示されている方法を適用する。上記認識対象語彙として、認識対象語彙セットＡと認識対象語彙セットＢとがあり、現時点においては認識対象語彙識別子記憶部７には認識対象語彙セットＡの識別子が記憶されているものとする。また、出力部１３は、認識対象語彙セットＡに対応する表示内容を表示しているものとする。

【０００９】この状態で、所定時間が経過すると、タイマ部８から認識対象語彙切り換え要求部９及び出力部１３に対して通知がなされる。そうすると、出力部１３は、表示内容を認識対象語彙セットＢに対応する表示内容に変更する。また、認識対象語彙切り換え要求部９から切り換えが要求され、その要求時刻が認識対象語彙切り換え要求時刻記憶部１０に記憶される。そして、認識対象語彙格納･判定部６によって、認識対象語彙切り換え要求時刻記憶部１０に記憶されている要求時刻Ｔcと音声時刻記憶部１２に記憶されている音声入力開始時刻Ｔsとが比較され、音声入力開始時刻Ｔsが要求時刻Ｔcよりも後である場合には、認識対象語彙の切り換えが要求された後に発声が行われたのであるから適切な認識対象語彙セットは認識対象語彙セットＢであると判定される。それ以外は、認識対象語彙セットＡであると判定される。そして、該当する認識対象語彙セットの識別子で現認識対象語彙識別子記憶部７の記憶内容を更新するのである。

【００１０】こうして、適切な認識対象語彙セットの判定が終了すると、認識部４は、音響分析部３で得られた特徴ベクトルと、現認識対象語彙識別子記憶部７に記憶されている識別子に対応して認識対象語彙格納･判定部６から出力される何れかの認識対象語彙セットを構成する各単語の音素列と、音響モデル格納部５に格納されているＨＭＭを用いて、以下のようにして音声認識を行う。

【００１１】すなわち、先ず、上記認識対象語彙に含まれる各単語のＨＭＭを求める。具体的には、音響モデル格納部５に記憶されている各音素のＨＭＭを、認識対象語彙セットを構成している各単語の音素列に対応させて結合するのである。

【００１２】次に、夫々の単語のＨＭＭについて、音響分析部３からの特徴ベクトルを用いて生起確率を求める。ＨＭＭによる音声認識においては、音声は初期状態から最終状態までの状態遷移の間にＨＭＭから出力されるシンボルの時系列として表される。そこで、初期状態の確率を任意の値に定め、順次状態遷移毎に出力確率および遷移確率を掛けていくことによって、発声がそのモデルＭ(単語のＨＭＭ)から発生される確率を求めることができる。逆に、発声を観測した場合に、その発声があるモデルＭから発生したと仮定すると、そのモデルＭからの発生の確率が計算できることになる。

【００１３】以下、上記認識部４における認識アルゴリズムについて詳細に説明する。認識部４は、音響分析部３によって得られた特徴ベクトルの時系列を入力とし、認識対象語彙格納･判定部６からの認識対象語彙に含まれる総ての単語のＨＭＭに関してその生起確率を求め、最も高い生起確率を呈するＨＭＭの単語を認識結果とする。すなわち、ｔ(＝１,２,…,Ｉ)をフレーム番号として、特徴ベクトルの時系列で表現された入力の系列を、

$X = x_{vec}1, x_{vec}2, x_{vec}3, \cdots, x_{vec}t, \cdots, x_{vec}I$

とする。尚、「$x_{vec}i$」は多次元のベクトルである。以下、ベクトルｘを「x_{vec}」と表記する。さらに、モデルＭの初期状態の集合をＳとし、最終状態の集合をＦとする。また、「ｉ，ｊ」を状態番号として、ｊ番目の状態の遷移系列を

$Q = q_0^j, q_1^j, q_2^j, \cdots, q_t^j, \cdots, q_I^j$

と表す。上式において、「q_t^j」は、ｔ番目のフレームの入力記号$x_{vec}t$によって遷移した状態を表す。ここで、$q_0^j \in S$であり、$q_I^j \in F$である。更に、初期状態の初期確率をπ_i：$\Sigma_{q_i \in S}\pi_i = 1$で表し、状態$q_i$から状態$q_j$への遷移確率を$a_{ij}$とし、そのときに$x_{vec}i$が出力される出力確率を$b_{ij}(x_{vec}i)$とすると、入力系列の生起確率(尤度)Ｐ(Ｘ|Ｍ)は、

$$P(X|M) = \sum_{all Q^j} \pi_0^j \prod_{i=1}^{I} a_{i-1,i}^j \cdot b_{i-1,i}^j (x_{vec} i)$$

で表される。この生起確率(尤度)Ｐ(Ｘ|Ｍ)の演算を、認識対象語彙に含まれる全単語に対応するＨＭＭついて計算し、最も高い生起確率(尤度)Ｐを呈するＨＭＭに対応する単語を認識結果として出力部１３に出力して表示するのである。

【００１４】

【発明が解決しようとする課題】しかしながら、上記従来の特開平６－３３７６９５号公報に開示された認識対象語彙切り換え動作を適用した音声認識装置には、以下のような問題がある。すなわち、上述したように、特開平６－３３７６９５号公報に開示された認識対象語彙切り換え動作においては、音声入力開始時刻Ｔｓが認識対象語彙切り換え要求時刻Ｔｃよりも後である場合に認識対象語彙のセットを切り換えるようにしている。この方法は、話者の操作によって認識対象語彙切り換え要求がなされる場合には、必ず認識対象語彙の切り換え要求がなされた後に発声が行われるために有効である。

【００１５】ところが、図４に示す音声認識装置のように、時間の経過と共に自動的に認識対象語彙が切り換る音声認識装置の場合には、認識対象語彙の切り換えは、話者の意識とは全く関係なく行われる。したがって、何らかの理由で話者が認識対象語彙の発声の機会を逸してしまい、且つ、自動的に認識対象語彙の切り換えが行われた場合には、何らかの方法によって話者が発声したかった切り換え前の認識対象語彙の設定状態にもどす必要が生ずる。そして、その場合には、何らかの操作を話者に負担させるか、若しくは、自動的に切り換え前の認識対象語彙が設定されるまで話者を待たせることになるという問題がある。

【００１６】そこで、この発明の目的は、自動的に認識対象語彙を切り換える場合でも高い認識精度が得られる使い易い音声認識装置および音声認識方法、並びに、音声認識処理プログラムを記録したプログラム記録媒体を提供することにある。

【００１７】

【課題を解決するための手段】上記目的を達成するため、第１の発明は、入力された音声を認識する認識部と,この認識部の認識結果を含む情報を出力する出力部と,上記認識時に用いられる認識対象語彙が格納された認識語彙格納部と,タイマ部と,このタイマ部からの時刻信号に基づいて上記認識対象語彙の切り換えを要求する認識対象語彙切り換え要求部を有する音声認識装置において、上記出力部は,複数の出力内容を切り換え出力するようになっており、上記認識対象語彙は,上記出力部の出力内容に対応した認識対象語の集合でなる複数の認識対象語彙セットに分類され,上記認識対象語彙の切り換えは上記認識対象語彙セットの単位で行われるようになっており、上記タイマ部からの時刻信号に基づいて,上記各認識対象語彙セット用の重みを決定する重み決定部を備えて、上記認識部は,上記全認識対象語彙セットおよび上記決定された各重みを用いて,入力音声を認識するようになっていることを特徴としている。

【００１８】上記構成によれば、認識部によって、全認識対象語彙セットおよびタイマ部からの時刻信号に基づいて重み決定部によって決定された各認識対象語彙セット用の重みを用いて、入力音声が認識される。その際に、上記タイマ部からの時刻信号に基づいて認識対象語彙切り換え要求部によって認識対象語彙の切り換えが要求されると、現在用いられている認識対象語彙セットが、出力部の出力内容の切り換えに応じた認識対象語彙セットに切り換えられる。したがって、切り換え前の認識対象語彙セット用の重みの値を低めるようにすれば、上記出力部の出力内容に対応している切り換え後の認識対象語彙の認識精度が高められる。

【００１９】さらに、話者が、上記認識対象語彙セットが切り換えられたことを知らずに、切り換え前の認識対象語彙で発声したとしても、切り換え前の認識対象語彙セットの語をも用いて認識が行われているので、上記切り換え前の認識対象語彙セットの語に関しても高い認識精度が得られる。

【００２０】また、上記第１の発明の音声認識装置は、上記重み決定部を、上記認識対象語彙切り換え要求部によって認識対象語彙の切り換えが要求されてから重み決定までの経過時間に応じて、切り換え前の認識対象語彙セット用の重みを低下させる一方、切り換え後の認識対象語彙セット用の重みを上昇させるように成すことが望ましい。

【００２１】上記構成によれば、上記認識対象語彙切り換え要求部によって認識対象語彙の切り換えが要求されてからの経過時間が長くなるに連れて、切り換え前の認識対象語彙の認識精度が低くなる一方、切り換え後の認識対象語彙の認識精度が高くなる。こうして、認識に用いられる上記認識対象語彙の切り換えが徐々に行われる。

【００２２】また、上記第１の発明の音声認識装置は、

上記認識部を、上記全認識対象語彙セットを構成する各語の尤度を算出し、各語の尤度の値に各語が属する認識対象語彙セット用の重みを掛け、その値が最も高い語を認識結果とするように成すことが望ましい。

【００２３】上記構成によれば、認識に用いられている認識対象語彙セット用の重みと認識に用いられていない認識対象語彙セット用の重みとを最適に設定することによって、上記出力部の出力内容に対応した切り換え後の認識対象語彙の認識精度を高めることと、話者が切り換え前の認識対象語彙で発声した場合でも高い認識精度を得ることとが、容易に達成される。

【００２４】また、上記第１の発明の音声認識装置は、上記出力部を、上記認識対象語彙切り換え要求部からの認識対象語彙切り換え要求がなされた時点に出力している出力内容に対応する認識対象語彙セット用の重みの値と、次に出力すべき出力内容に対応する認識対象語彙セット用の重みの値との差が所定値未満になると、上記出力内容を切り換えるように成すことが望ましい。

【００２５】上記構成によれば、上記認識対象語彙セットが切り換えられるのに呼応して、上記出力部の出力内容が対応する出力内容に切り換えられる。

【００２６】また、第２の発明の音声認識方法は、入力された音声を認識対象語彙を用いて認識して認識結果を出力するに際して,タイマ部からの時刻信号に基づいて上記認識対象語彙の切り換えを自動的に行う音声認識方法において、複数の出力内容を出力部に切り換え出力し、上記各出力内容に対応した認識対象語の集合でなる複数の認識対象語彙セットの単位で上記認識対象語彙の切り換えを行い、上記タイマ部からの時刻信号に基づいて上記各認識対象語彙セット用の重みを決定し、上記全認識対象語彙セットおよび上記決定された各重みを用いて上記入力音声の認識を行うことを特徴としている。

【００２７】上記構成によれば、全認識対象語彙セットおよびタイマ部からの時刻信号に基づいて決定された各認識対象語彙セット用の重みを用いて、入力音声が認識される。その際に、上記タイマ部からの時刻信号に基づいて認識対象語彙の切り換えが要求されると、現在用いられている認識対象語彙セットが、出力部の出力内容の切り換えに応じた認識対象語彙セットに切り換えられる。したがって、切り換え前の認識対象語彙セット用の重みの値を低めるようにすれば、上記出力部の出力内容に対応している切り換え後の認識対象語彙の認識精度が高められる。

【００２８】さらに、話者が、上記認識対象語彙セットが切り換えられたことを知らずに、切り換え前の認識対象語彙で発声したとしても、切り換え前の認識対象語彙セットの語をも用いて認識が行われているので、上記切り換え前の認識対象語彙セットの語に関しても高い認識精度が得られる。

【００２９】また、第３の発明のプログラム記録媒体は、コンピュータを、請求項１における認識部,出力部,タイマ部,認識対象語彙切り換え要求部および重み決定部として機能させる音声認識処理プログラムが記録されていることを特徴としている。

【００３０】上記構成によれば、請求項１の場合と同様に、切り換え前の認識対象語彙セット用の重みの値を低めるようにすれば、上記出力部の出力内容に対応している切り換え後の認識対象語彙の認識精度が高められる。さらに、話者が、上記認識対象語彙セットが切り換えられたことを知らずに切り換え前の認識対象語彙で発声したとしても、高い認識精度が得られる。

【００３１】

【発明の実施の形態】以下、この発明を図示の実施の形態により詳細に説明する。図１は、本実施の形態の音声認識装置におけるブロック図である。この音声認識装置２１は、音声入力部２２,Ａ/Ｄ変換部２３,音響分析部２４,認識部２５,音響モデル格納部２６,第１認識対象語彙セット格納部２７,第２認識対象語彙セット格納部２８,重み係数決定部２９,タイマ部３０,認識対象語彙切り換え要求部３１および出力部３２で構成される。

【００３２】上記音声入力部２２は、マイクロホンを含む音声入力装置を備えて、入力された音声を電気信号(音声信号)に変換してＡ/Ｄ変換部２３に出力する。Ａ/Ｄ変換部２３は、入力されたアナログ信号である音声信号をディジタル信号に変換し、ディジタル化された音声信号を音響分析部２４に出力する。尚、上記ディジタル化された音声信号は、振幅値の時系列で表されている。

【００３３】上記音響分析部２４は、Ａ/Ｄ変換部２３からのディジタル音声信号からフレーム毎に特徴ベクトルを抽出して認識部２５に出力する。ここで、上記特徴ベクトルは、各フレームにおける音声信号のパワー,１次～１６次のＬＰＣケプストラム係数,前フレームのパワーおよび前フレームのＬＰＣケプストラム係数(１次～１６次)の合計３４の要素からなる３４次元ベクトル$x_{t,i}$を、総てのフレーム（ｔ＝１,２,…,Ｉ)に亘って配列したものである。

【００３４】上記認識部２５は、音響モデルを利用して、音響分析部２４で抽出された特徴ベクトルを用いて、第１認識対象語彙セット格納部２７に格納されている認識対象語彙セットＡおよび第２認識対象語彙セット格納部２８に格納されている認識対象語彙セットＢを構成する各単語の生起確率(尤度)Ｐを、従来の技術で説明した手法を用いて計算する。さらに、重み係数決定部２９で決定された重みｗを各単語の尤度Ｐに掛けて、最も高い尤度ｗ･Ｐを呈するＨＭＭに対応する単語を出力部３２に出力するのである。

【００３５】上記音響モデル格納部２６は、認識部２５で音声認識を行う際に使用される音響モデルが格納されている。上記音響モデルは、音素を単位として、予め不特定話者の学習音声を用いてバウム・ウェルチアルゴリ

ズムと呼ばれるアルゴリズムによって学習(初期学習)されたHMMが用いられる。尚、上記HMMは、各状態における遷移確率と出力確率分布とを要素とする状態数分の配列で記憶されている。また、上記遷移確率は、各状態への遷移確率を要素として遷移数分の配列で記憶されている。また、上記出力確率は、複数の正規分布を重み付け加算した多次元の混合正規分布で表され、各正規分布における混合の重みと平均ベクトルと分散ベクトルとを要素とする次元数分の配列で記憶されている。ここで、上記平均ベクトルと分散ベクトルとは、音響分析部２４で音声信号からフレーム毎に抽出される特徴ベクトルの要素数と同じ「３４」の要素の配列で表される。

【００３６】上記タイマ部３０は、時刻を表す時刻信号を認識対象語彙切り換え要求部３１,重み係数決定部２９および出力部３２に出力して、時刻を通知する。そうすると、認識対象語彙切り換え要求部３１は、上記通知された時刻に基づいて、認識対象語彙の切り換えを要求するか否かを判断する。そして、要求する場合には、重み係数決定部２９に対して認識対象語彙の切り換えを要求する。

【００３７】上記重み係数決定部２９は、第１認識対象語彙セット格納部２７に格納されている認識対象語彙セットＡおよび第２認識対象語彙セット格納部２８に格納されている認識対象語彙セットＢのうち、出力部３２によって現在表示されている表示内容に対応する認識対象語彙セットを構成する単語に掛けられる重みw_1、および、出力部３２によって表示されていない表示内容に対応する認識対象語彙セットを構成する単語に掛けられる重みw_2を決定する。これらの重みw_1,w_2は、記憶されている重み関数$W_1(t)$,$W_2(t)$を用いて、認識対象語彙切り換え要求部３１から切り換えが要求された時のタイマ部３０からの時刻t_0を基準として所定時間△Ｔが経過する毎に決定される。そして、決定された両重みw_1,w_2の値は認識部２５に順次出力される。

【００３８】上記第１認識対象語彙セット格納部２７および第２認識対象語彙セット格納部２８には、夫々の認識対象語彙セットＡ,Ｂを構成する単語が、各単語の表記と音素列との文字列を要素とする文字数分の配列で記憶されている。

【００３９】上記出力部３２は、ディスプレイを含む画像表示装置を備えて、認識対象語彙セットＡに対応した第１表示内容と認識対象語彙セットＢに対応した第２表示内容とを格納している。そして、タイマ部３０から通知される時刻に基づいて、第１,第２表示内容のうち現在表示している表示内容を変更するか否かを判断し、変更する場合は画面の表示内容を切り換える。さらに、認識部２５からの認識結果を画面に表示する。

【００４０】図２は、上記出力部３２が現在選択している表示内容に対応する認識対象語彙セット用の重み関数$W_2(t)$と非選択表示内容に対応する認識対象語彙セット用の重み関数$W_1(t)$との時間変化を示す。重み関数$W_1(t)$の値は、認識対象語彙の切り換え要求が出力された時刻t_0で１よりも小さい０近傍の所定値「ａ」から単調増加し始め、時刻t_2以降は値「１」となる。一方、重み関数$W_2(t)$の値は、重み関数$W_1(t)$の値とは逆に、時刻t_0で値「１」から単調減少し始めて、時刻t_2以降は所定値「ａ」となる。その場合、時刻t_1で重みw_1と重みw_2の差が閾値ｈとなる。そして、出力部３２は、この差の値が閾値ｈ未満になると、つまり認識対象語彙の切り換えが要求された時刻t_0から時間Ｔ($>(t_2-t_0)$)が経過すると、画面に表示されている表示内容を切り換えるのである。

【００４１】すなわち、上記出力部３２がタイマ部３０から通知される時刻に基づいて表示内容を変更すると判断する時点は、認識対象語彙切り換え要求部３１がタイマ部３０から通知された時刻に基づいて上記切り換えを要求すると判断する時点よりも上記時間Ｔだけ遅れるように設定されているのである。

【００４２】このように、本実施の形態においては、出力部３２によって、自動的に画面の表示内容が切り換えられるのであるが、表示内容が切り換る前であっても切り換った後であっても、認識部２５は、認識対象語彙セットＡおよび認識対象語彙セットＢの両語彙セットの語彙を対象として尤度Ｐの計算を行う。そして、現在出力部３２によって選択されている表示内容に対応する認識対象語彙セットを構成する単語の尤度Ｐには、表示内容切り換え前であれば$1>w>(1+a+h)/2$であり、切り換え後であれば$1>w>(1+a-h)/2$である重みｗを掛ける。一方、非選択側の表示内容に対応する認識対象語彙セットを構成する単語の尤度Ｐには、表示内容切り換え前であれば$(1+a-h)/2>w>a$であり、切り換え後であれば$(1+a+h)/2>w>a$である重みｗを掛ける。こうして、最終的な尤度ｗ･Ｐを計算して認識結果を決定するようにしている。

【００４３】換言すれば、図４に示す従来の音声認識装置における認識対象語彙の切り換えは、尤度Ｐの演算に用いる認識対象語彙そのものを切り換えることによって行うのに対して、本実施の形態においては、尤度Ｐの演算に用いる２セットの認識対象語彙は切り換えずに尤度Ｐに掛ける重みｗの値を「１」と０近傍の所定値「ａ」との間で徐々に変化させることによって行うのである。

【００４４】したがって、本実施の形態においては、何らかの理由で話者が認識対象語彙の発声の機会を逸してしまい、且つ、自動的に認識対象語彙の切り換えが行われた後でも、切り換え前の認識対象語彙の単語に関する尤度ｗ･Ｐの計算も行われることになり、話者が切り換え前の認識対象語彙で発声しても正しく認識することが可能になる。また、その場合、図４に示す音声認識装置のように認識対象語彙そのものを切り換えた場合と同様に、出力部３２の表示内容に対応した語彙の認識精度を

高める機能は損なわれないのである。

【００４５】図３は、上記重み係数決定部２９によって実行される重み決定処理動作のフローチャートである。以下、図３に従って、重み決定の動作について説明する。ここで、出力部３２が現在選択している表示内容に対応する認識対象語彙セット用の重み関数を$W_2(t)$とし、非選択表示内容に対応する認識対象語彙セット用の重み関数を$W_1(t)$とする。認識対象語彙切り換え要求部３１から切り換えが要求されると重み決定処理動作がスタートする。

【００４６】ステップＳ1で、上記タイマ部３０からの時刻信号に基づいて、認識対象語彙の切り換え要求時刻t_0が取得される。ステップＳ2で、重み値wの算出回数jが「０」に初期化される。ステップＳ3で、算出回数jがインクリメントされる。ステップＳ4で、切り換え要求時刻t_0を取得してから又は前回重み値wを算出してから所定時間ΔＴが経過したか否かが判別される。その結果、経過していればステップＳ5に進む。ステップＳ5で、現在の時刻($t_0+j\cdot\Delta T$)が時刻t_2を越えているか否かが判別される。その結果、超えていれば重み決定処理動作を終了する一方、越えていなければステップＳ6に進む。

【００４７】ステップＳ6で、上記重み関数$W_i(t)$の関数番号iが「１」に初期化される。ステップＳ7で、重み関数$W_1(t)$における切り換え要求時刻t_0からの経過時間tに「$j\cdot\Delta T$」が代入されて、重みの値w_1が算出される。ステップＳ8で、関数番号iがインクリメントされる。ステップＳ9で、関数番号iの値が「２」よりも大きいか否かが判別される。その結果、「２」以下であればステップＳ7にリターンして重み値w_2の算出に移行する一方、「２」よりも大きければ、総ての認識対象語彙セットＡ,Ｂに対応する現時刻での重みが算出されたと判断されて、ステップＳ10に進む。ステップＳ10で、上記算出された現時刻での重み値w_1,w_2の配列が認識部２５に出力される。そうした後、ステップＳ3にリターンして、次の時刻での重み値w_1,w_2の算出に移行する。

【００４８】以後、上記ステップＳ3～ステップＳ10を繰り返し、ステップＳ5において現在の時刻が時刻t_2を越えていると判別されると、重み決定処理動作を終了する。その後は、表示内容に対応する認識対象語彙セット用の重み値w_2として「１」が所定時間ΔＴ毎に出力され、非選択表示内容に対応する認識対象語彙セット用の重み値w_1として所定値「ａ」が所定時間ΔＴ毎に出力される。そして、次に認識対象語彙切り換え要求部３１から切り換え要求が出力されると、上記重み決定処理動作がスタートするのである。

【００４９】上述のように、本実施の形態における認識部２５は、音響モデル格納部２６に格納された音響モデルを用いて、第１認識対象語彙セット格納部２７に格納された認識対象語彙セットＡと第２認識対象語彙セット格納部２８に格納された認識対象語彙セットＢとを構成する単語の尤度Ｐを算出する。その際における出力部３２の表示内容の切り換えに伴う認識対象語彙セットの切り換えは、認識対象語彙セットそのものを切り換えるのではなく、選択,非選択認識対象語彙セットを構成する単語の尤度Ｐに掛ける重みw_2,w_1の値を「１」と０近傍の所定値「ａ」とに切り換えることによって行う。そして、その場合に、重みw_2,w_1の値を段階的に切り換えるのではなく、認識対象語彙切り換え要求部３１から切り換え要求がなされた時刻t_0からの経過時間「$j\cdot\Delta T$」に比例して徐々に値「１」から値「ａ」へ又は値「ａ」から値「１」へ切り換えるようにしている。

【００５０】したがって、本実施の形態によれば、何らかの理由で話者が認識対象語彙の発声の機会を逸してしまい、且つ、自動的に認識対象語彙が切り換えられてしまっても、切り換え前の認識対象語彙セットの単語に関する尤度$w\cdot P$の計算をも行うので、話者が切り換え前の認識対象語彙で発声しても正しく認識することができる。また、その場合に、図４に示す音声認識装置のごとく認識対象語彙そのものを切り換える場合と同様に、出力部３２の表示内容に対応した認識対象語彙の認識精度を高める機能は損なわれることはない。

【００５１】尚、上記実施の形態においては、選択認識対象語彙セット用の重み関数$W_2(t)$および非選択表示内容に対応する認識対象語彙セット用の重み関数$W_1(t)$を、認識対象語彙切り換え要求部３１による切り換え要求時刻t_0からの経過時間「$j\cdot\Delta T$」に比例して、値「１」,「ａ」から値「ａ」,「１」へ直線的に切り換えるようにしている。しかしながら、この発明においては、関数$W_1(t)$,$W_2(t)$の形状は直線に限定されるものではない。曲線にして、表示内容の切り換え時刻t_1までの関数$W_2(t)$の値を高める一方関数$W_1(t)$の値を低め、表示内容の切り換え時刻t_1以降の関数$W_2(t)$の値を低める一方関数$W_1(t)$の値を高めてもよい。

【００５２】また、上記実施の形態においては、上記重み係数決定部２９を、認識対象語彙切り換え要求部３１からの切り換え要求時刻t_0を基準として所定時間ΔＴが経過する毎に重み値w_1,w_2を決定して認識部２５に出力するように構成し、認識部２５は、入力される重み値w_1,w_2を必要に応じて用いて認識処理を行うように構成している。しかしながら、この発明はこれに限定されるものではなく、認識部２５を、認識を行う際に重み係数決定部２９に対して重み決定要求を出すように構成し、重み係数決定部２９は、重み決定要求を受けると、認識対象語彙切り換え要求部３１による切り換え要求時刻t_0からの経過時間を重み関数$W_i(t)$に代入して算出するように構成しても差し支えない。

【００５３】ところで、上記各実施の形態における上記認識部,出力部,タイマ部,認識対象語彙切り換え要求部および重み決定部としての機能は、プログラム記録媒体

に記録された音声認識処理プログラムによって実現される。上記実施の形態における上記プログラム記録媒体は、ＲＯＭ(リード･オンリ･メモリ)でなるプログラムメディアである。あるいは、外部補助記憶装置に装着されて読み出されるプログラムメディアであってもよい。尚、何れの場合においても、上記プログラムメディアから音声認識処理プログラムを読み出すプログラム読み出し手段は、上記プログラムメディアに直接アクセスして読み出す構成を有していてもよいし、ＲＡＭ(ランダム･アクセス･メモリ)に設けられたプログラム記憶エリア(図示せず)にダウンロードし、上記プログラム記憶エリアにアクセスして読み出す構成を有していてもよい。尚、上記プログラムメディアからＲＡＭの上記プログラム記憶エリアにダウンロードするためのダウンロードプログラムは、予め本体装置に格納されているものとする。

【００５４】ここで、上記プログラムメディアとは、本体側と分離可能に構成され、磁気テープやカセットテープ等のテープ系、フロッピー（登録商標）ディスク,ハードディスク等の磁気ディスクやＣＤ(コンパクトディスク)－ＲＯＭ,ＭＯ(光磁気)ディスク,ＭＤ(ミニディスク),ＤＶＤ(ディジタルビデオディスク)等の光ディスクのディスク系、ＩＣ(集積回路)カードや光カード等のカード系、マスクＲＯＭ,ＥＰＲＯＭ（紫外線消去型ＲＯＭ),ＥＥＰＲＯＭ(電気的消去型ＲＯＭ),フラッシュＲＯＭ等の半導体メモリ系を含めた、固定的にプログラムを担持する媒体である。

【００５５】また、上記各実施の形態における音声認識装置は、モデムを備えてインターネットを含む通信ネットワークと接続可能な構成を有していれば、上記プログラムメディアは、通信ネットワークからのダウンロード等によって流動的にプログラムを担持する媒体であっても差し支えない。尚、その場合における上記通信ネットワークからダウンロードするためのダウンロードプログラムは、予め本体装置に格納されているものとする。あるいは、別の記録媒体からインストールされるものとする。

【００５６】尚、上記記録媒体に記録されるものはプログラムのみに限定されるものではなく、データも記録することが可能である。

【００５７】

【発明の効果】以上より明らかなように、第１の発明の音声認識装置は、出力部の出力内容に対応した複数の認識対象語彙セットを認識語彙格納部に格納し、重み決定部によって、タイマ部からの時刻信号に基づいて上記各認識対象語彙セット用の重みを決定し、認識部によって、上記全認識対象語彙セットおよび上記決定された各重みを用いて入力音声を認識するので、認識対象語彙切り換え要求部による認識対象語彙の切り換え要求に基づいて、上記出力部の出力内容の切り換えに応じた認識対象語彙セットに切り換えられる際に、切り換え前の認識対象語彙セット用の重みの値を低めるようにすれば、切り換え後の認識対象語彙の認識精度を高めることができる。

【００５８】さらに、話者が、上記認識対象語彙セットが切り換えられたことを知らずに、切り換え前の認識対象語彙で発声しても、切り換え前の認識対象語彙セットの語をも用いて認識を行うので、上記切り換え前の認識対象語彙セットの語に関しても高い認識精度を得ることができる。

【００５９】すなわち、この発明によれば、自動的に認識対象語彙を切り換える場合でも高い認識精度を得ることができる。さらに、その際に話者に何らかの操作や待ち時間を負担させることがなく、使い易い音声認識装置を実現できる。

【００６０】また、上記第１の発明の音声認識装置は、上記重み決定部を、上記認識対象語彙切り換え要求部によって認識対象語彙の切り換えが要求されてから重み決定までの経過時間に応じて、切り換え前の認識対象語彙セット用の重みを低下させる一方、切り換え後の認識対象語彙セット用の重みを上昇させるように成せば、認識に用いられる上記認識対象語彙の切り換えを徐々に行うことができる。したがって、上記切り換え前の認識対象語彙セットの語に関しても高い認識精度を得ることができる。

【００６１】また、上記第１の発明の音声認識装置は、上記認識部を、全認識対象語彙セットを構成する各語の尤度を算出し、各語の尤度の値に各語が属する認識対象語彙セット用の重みを掛け、その値が最も高い語を認識結果とするように成せば、認識に用いられている認識対象語彙セット用の重みと認識に用いられていない認識対象語彙セット用の重みとを最適に設定すれば、上記出力部の出力内容に対応している切り換え後の認識対象語彙の認識精度を高めることと、話者が切り換え前の認識対象語彙で発声した場合でも高い認識精度を得ることとを、容易に達成することができる。

【００６２】また、上記第１の発明の音声認識装置は、上記出力部を、上記認識対象語彙切り換え要求部からの認識対象語彙切り換え要求がなされた時点に出力している出力内容に対応する認識対象語彙セット用の重みの値と、次に出力すべき出力内容に対応する認識対象語彙セット用の重みの値との差が所定値未満になると、上記出力内容を切り換えるように成せば、上記認識対象語彙セットが切り換えられるのに呼応して、上記出力部の出力内容を対応する出力内容に切り換えることができる。

【００６３】また、第２の発明の音声認識方法は、タイマ部からの時刻信号に基づいて出力部の出力内容に対応した複数の認識対象語彙セット用の重みを決定し、全認識対象語彙セットおよび上記決定された各重みを用いて入力音声を認識するので、認識対象語彙セットが切り換

えられる際に、切り換え前の認識対象語彙セット用の重みの値を低めるようにすれば、上記出力部の出力内容に応じた切り換え後の認識対象語彙の認識精度を高めることができる。

【００６４】さらに、話者が、上記認識対象語彙セットが切り換えられたことを知らずに、切り換え前の認識対象語彙で発声しても、切り換え前の認識対象語彙セットの語をも用いて認識を行うので、上記切り換え前の認識対象語彙セットの語に関しても高い認識精度を得ることができる。

【００６５】また、第３の発明のプログラム記録媒体は、コンピュータを、請求項１における認識部,出力部,タイマ部,認識対象語彙切り換え要求部および重み決定部として機能させる音声認識処理プログラムが記録されているので、請求項１の場合と同様に、切り換え前の認識対象語彙セット用の重みの値を低めるようにすれば、上記出力部の出力内容に対応している切り換え後の認識対象語彙の認識精度を高めることができる。さらに、話者が、上記認識対象語彙セットが切り換えられたことを知らずに切り換え前の認識対象語彙で発声したとしても、高い認識精度を得ることができる。

【図面の簡単な説明】

＊【図１】　この発明の音声認識装置におけるブロック図である。

【図２】　選択,非選択認識対象語彙セット用の重み関数の時間変化を示す図である。

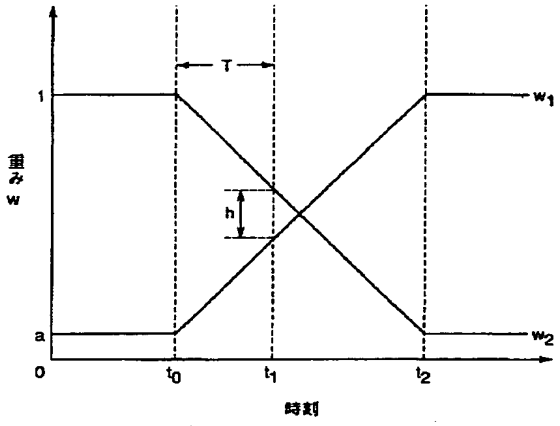
【図３】　図１における重み係数決定部によって実行される重み決定処理動作のフローチャートである。

【図４】　認識対象語彙の切り換えが可能な従来の音声認識装置のブロック図である。

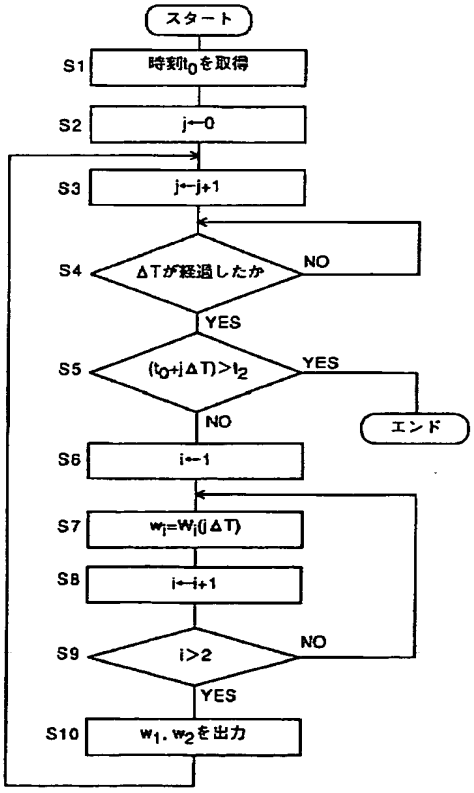
【符号の説明】

２１…音声認識装置、
２２…音声入力部、
２３…Ａ/Ｄ変換部、
２４…音響分析部、
２５…認識部、
２６…音響モデル格納部、
２７…第１認識対象語彙セット格納部、
２８…第２認識対象語彙セット格納部、
２９…重み係数決定部、
３０…タイマ部、
３１…認識対象語彙切り換え要求部、
３２…出力部。

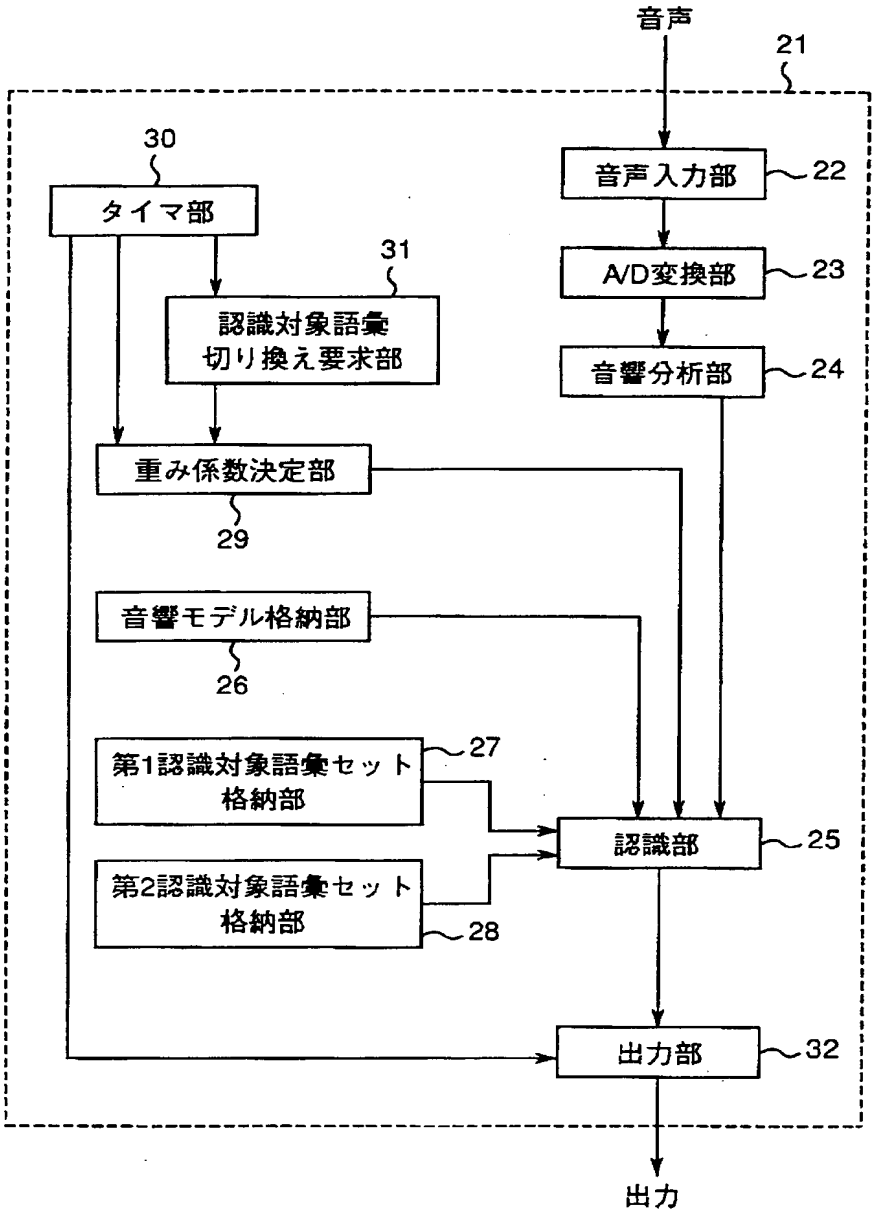
【図２】



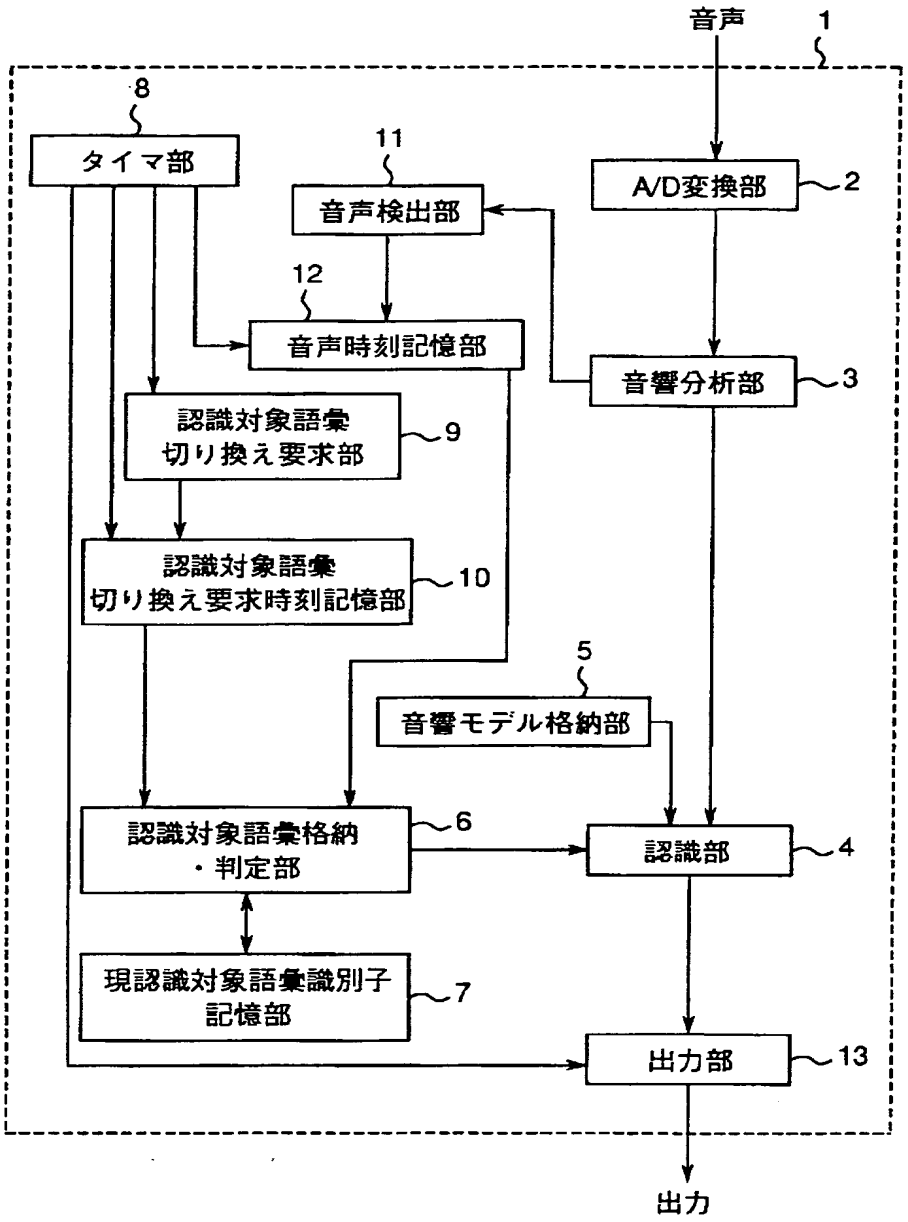
【図３】



【図１】


音声
21
30
タイマ部
31
認識対象語彙
切り換え要求部
音声入力部
22
A/D変換部
23
音響分析部
24
重み係数決定部
29
音響モデル格納部
26
第1認識対象語彙セット
格納部
27
第2認識対象語彙セット
格納部
28
認識部
25
出力部
32
出力

【図４】



フロントページの続き

(72)発明者　勘座　浩幸
大阪府大阪市阿倍野区長池町22番22号　シャープ株式会社内

Ｆターム(参考)　5D015 AA04 BB02 HH05 HH11 HH21 KK02 LL10